Zinc: Leveraging GPUs to Improve Simulation Speed

Preston Hansen, Elgin Community College & University of Illinois at Chicago - SULI Program
Mentor: Taritree Wongjirad

Improving the Existing Event Loop

- Want to generate and propagate photons quickly.
  - Liquid argon yields 10,000s of photons per MeV (~650k per event in RAT/Chroma!).
- Need a multiplatform, portable solution.
- Need a model that is modular and easily configurable.
- Need to minimize network impact (small messages!).

ZMQ
Easy-to-use socket API enabling cross-platform message passing

Protobuf
Efficient, multiplatform message serialization and parsing libraries

→ **Zinc**
Powerful, scalable networking layers to interface with RAT, Chroma, and other simulation tools

Why Use RAT/Chroma?

- RAT serves as an easily customizable development tool to create a model which can be ported to LArSoft.
  - Components of the RAT simulation can be toggled for easier debugging/prototyping.
  - RAT outputs ROOT files, which can be easily analyzed, even by those without experience with RAT.
- Chroma utilizes GPU (graphics processing unit) parallel processing capabilities to increase photon generation and propagation speeds.

RAT – Zinc – Chroma Event Loop

RAT
- Join Zinc queue.
- Send detector config. / geometry.
- Generate and send step-data.

- Store returned photon data into a GLG4HitPhoton vector.
- Output results to a ROOT file.

Zinc
- Queue requests.
- Pass data to Chroma.

- Identify correct client to return to.
- Efficiently pass new (serialized) message through server/queue.

The Zinc interface can be scaled to handle many clients simultaneously, thanks to the powerful ZMQ socket API powering message transfer!

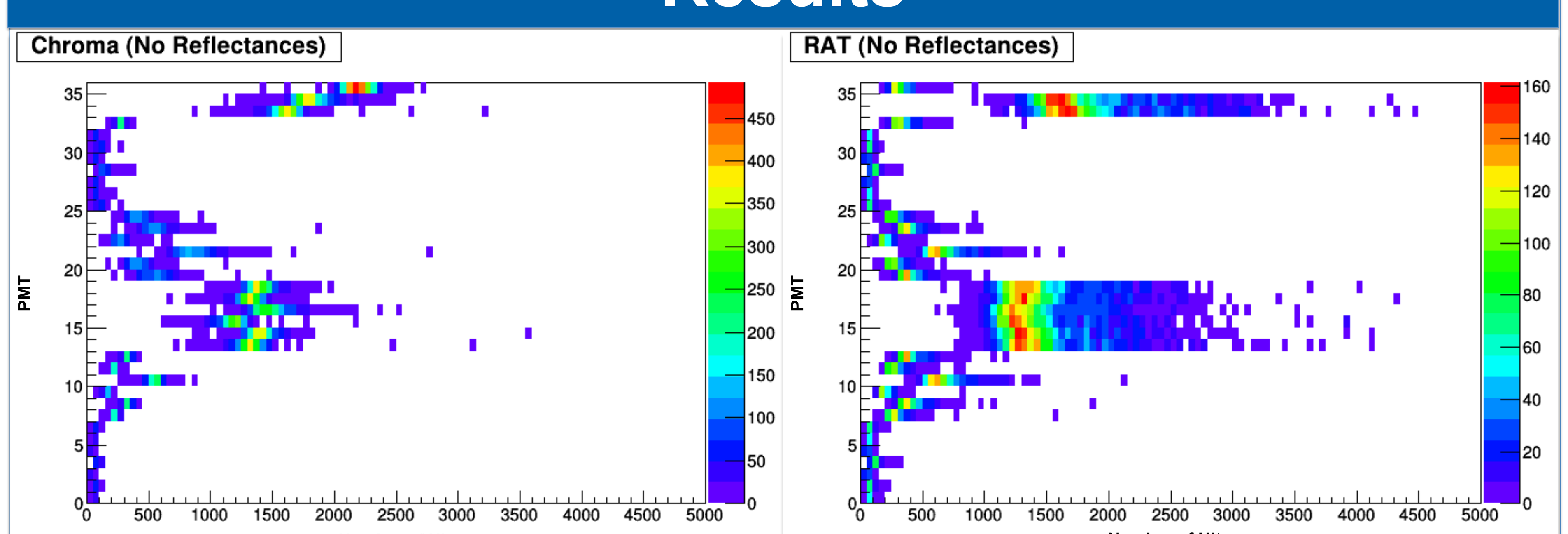
Chroma
- Generate photons from step-data.
- Simulate events in detector.
- Pass to native C++ layer for fast message creation/serialization.

- Return new data through Zinc.
- Wait for new step-data.

ZMQ/Protobuf

*Chroma allows for round-trip performance **4x to 8x faster** than CPU-based implementations (Nvidia K20 vs. AMD Phenom II x4 955).*

Results



Above: Number of hits vs. PMT for Chroma and RAT. Note that hit count for PMTs 32–35 is higher, as 32–35 are light paddles located near the center of the PMT array. (See figure below.)

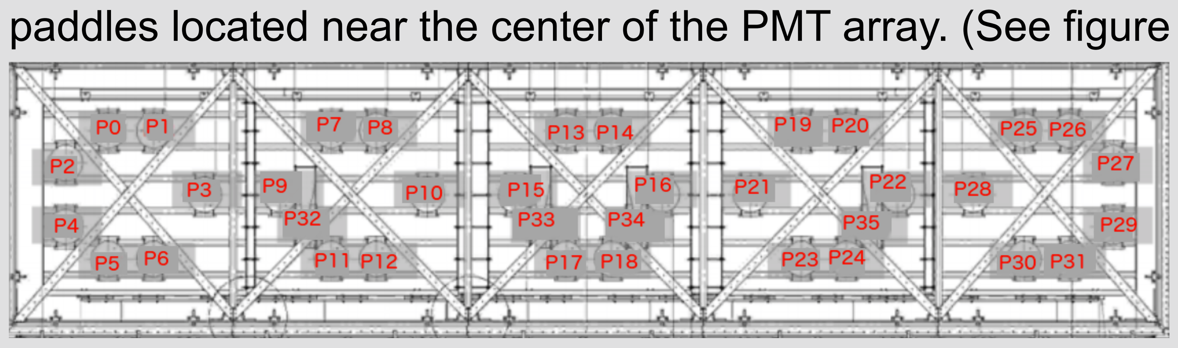
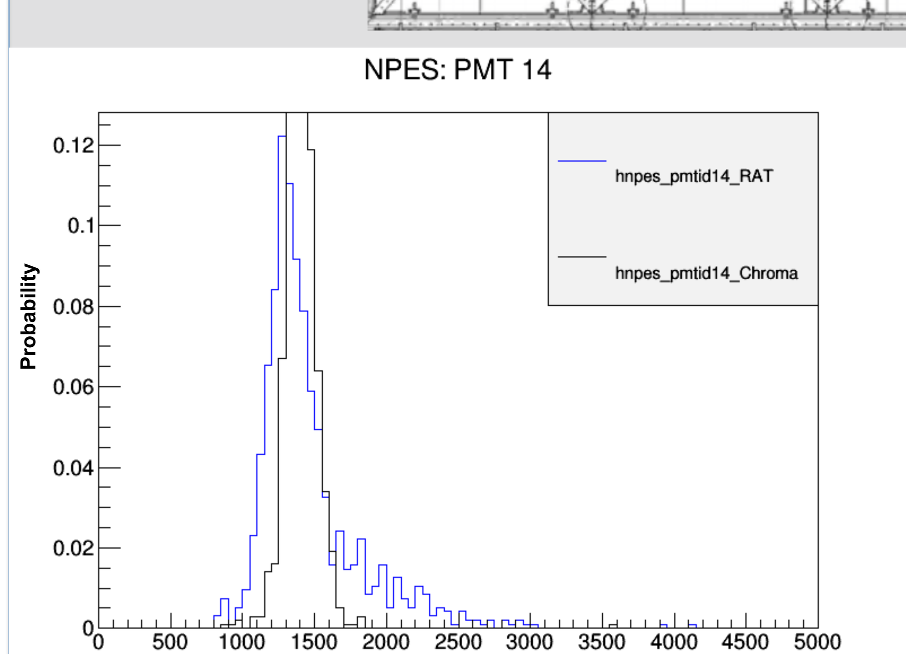
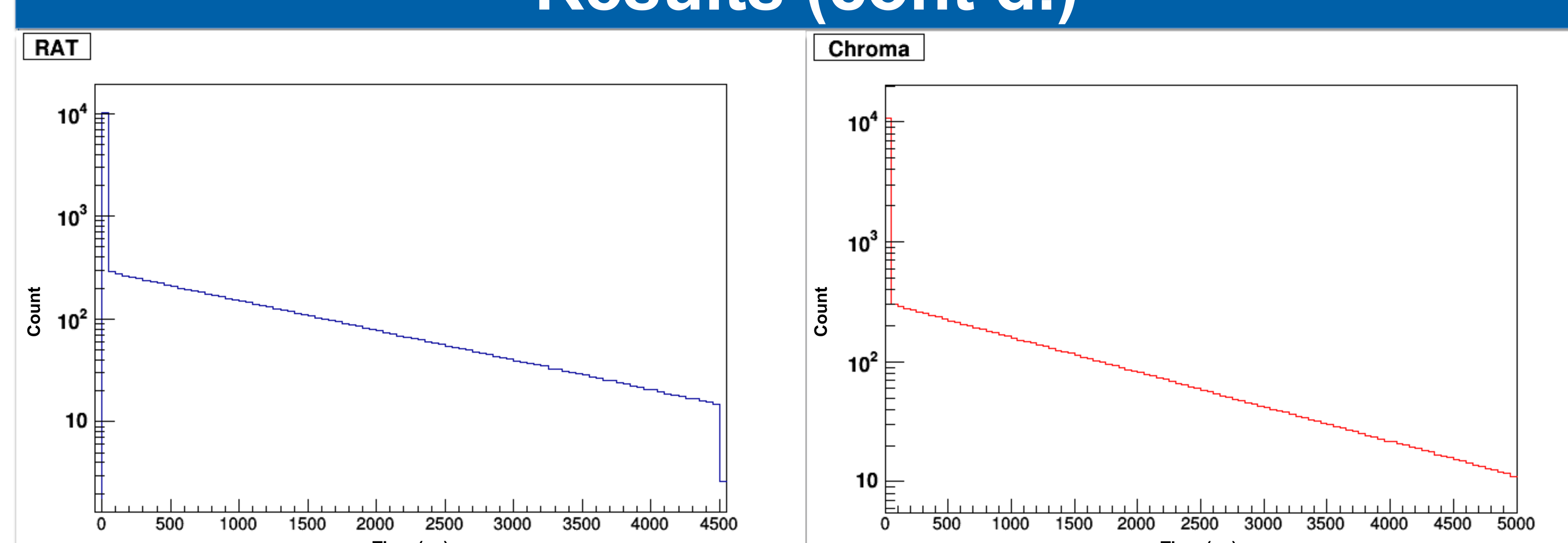


Image source: M. Toups, 8/9/13 MicroBooNE Collab. Meeting



Left: Probability of a hit count for a given PMT in RAT and Chroma. Note that although the means are similar, variance is markedly higher for RAT than Chroma. (This can be seen across all PMTs in the graphs above.)

Results (cont'd.)



Above: Number of hits vs. time (in ns) for RAT vs. Chroma. Both tools allowed modeling of prompt and late scintillator decay. Note that times in RAT past 4500 ns were not logged, reflected by the drop-off seen above.

Conclusions and Looking Forward

- GPU-based photon simulation improved round-trip performance between 4x (Nudot, MIT) and 8x (Oppenheimer, FNAL).
  - **Round-trip times reduced from 12–24 seconds to just 3 seconds**!
- Zinc supports many client requests asynchronously.
  - In conjunction with Geant 4.10+, could process multiple events simultaneously.
- Simulation could be scaled to incorporate multiple GPUs, even across multiple different machines.
- The model was successfully ported to LArSoft and could be further implemented to interface with LArSoft simulations.

August 2015

Elgin Community College · UIC · Fermilab · U.S. DEPARTMENT OF ENERGY · Office of Science